

REMARKS

The examiner has objected to claims 9 and 10 for certain informalities. Both of the issues raised by the examiner were addressed in an amendment filed on September 1, 2004. That amendment was apparently not before the examiner at the time that he worked on the office action. The amendment corrects the claim dependency of claim 9, and cancels claim 11 in view of its now being duplicative of claim 10.

The examiner has rejected claim 1 under 35 USC 112, paragraph 1, as failing to comply with the written description requirement. The limitation that the examiner asserts is not adequately supported in the original application is as follows:

relying on the data item already stored in the data repository for storage of the second data item rather than storing a separate copy of the second data item.

The examiner is urged to reconsider and withdraw the rejection, for this limitation is a central and prominent aspect of the invention. For example, the following paragraphs from the body of the application make it clear that only a single copy of a data item is kept in the data repository (except for copies made for data integrity purposes):

[0046] This repository is unusual in that, like the BBS systems cited above, from a logical standpoint it contains only a single copy of each data-item stored in it no matter how many repository clients (i.e., computers running software acting on behalf of human users) store files into it containing the same data-item. Any replication of data is done purely to assure data integrity (i.e., to make sure data is correct) and to improve data availability (i.e., to make sure a copy of the data is available) and accessibility (i.e., to make sure data can be accessed reasonably quickly). A pointer to a data-item already contained within this repository can be constructed directly from a copy of the same data-item present on a client computer, without the aid of the repository data-servers. Such pointers can be communicated to the repository in place of the actual data-items themselves.

[0047] The unusual organization of the repository is a key element in making significant network storage practicable for computers with slow network connections. Advantage is taken of the fact that most of the data on a typical computer duplicates data that is also present on other machines: operating system files, applications, and data files that have been downloaded over the network or

copied from removable media. In order to transfer such files to the repository, client software will typically only have to send a pointer, since the repository will already contain a copy of the data, sent earlier by some other client. An important element in the scheme is arranging to share data in this manner without compromising the privacy of user data--this is accomplished by sharing encrypted data.

* * *

[0058] The data repository is a distributed aggregate of data storage devices connected to the network, which together maintain a collection of data-items in a single logical address space, indexed by "datanames" (digital fingerprints) generated directly from the data-items themselves. Logically only one copy of each distinct data-item is kept in the repository, which allows for great economy in use of storage space. In practice, some redundancy is needed in order to assure data integrity, and to increase data availability and accessibility. Economical transmission of data-items which reside within the repository can be accomplished by sending the dataname in place of the data-item. This is illustrated in FIG. 1.

(Emphasis added)

The examiner has rejected claim 1, the only independent claim, under 35 USC 103(a) as being unpatentable over Shnelvar. The examiner is urged to reconsider and withdraw the rejection, particularly in light of the amendments made to claim 1. Shnelvar fails to teach or make obvious at least two features of claim 1.

The first of the features that Shnelvar fails to teach is the reliance on digital fingerprints to determine whether a data item is already stored in a data repository. Claim 1 has the following limitations:

having a second client program initiate a process for depositing a second data item in the data repository, wherein the second data item is identical to the data item stored by the first client program, the process including
determining a digital fingerprint from the second data item using the reproducible pseudorandom process; and
comparing the digital fingerprint from the second data item to digital fingerprints for data items already stored in the data repository, and determining from the comparing of digital fingerprints, without comparing the entire contents of the second data item to the entire contents of a data item already stored, that a

data item identical to the second data item is already stored in the data repository;
and

relying on the data item already stored in the data repository for storage of
the second data item rather than storing a separate copy of the second data item.

Shnelvar explicitly rejects the underscored feature of claim 1. Shnelvar does a
comparison of digital fingerprints (hash values) to determine whether the fingerprint of a new
data item ("new data unit") is the same as the fingerprint of an existing data item ("previously
stored data unit"), but Shnelvar refuses (e.g., col. 21, line 38 to col. 22, line 21) to rely on the
outcome of the fingerprint comparison, out of concern that there is a remote possibility that
different data items ("nonidentical data units") will produce identical fingerprints ("hash
values"). Shnelvar insists on doing a full comparison of the new data item with the existing data
item (col. 21, line 61 to col.22, line 8; emphasis added):

As such, the hashing algorithm and size of hash values are selected, for any given
implementation, so that the number of bits used in the hash values relative to the
number of allocation units is such that the likelihood that nonidentical data units
will result in identical hash values is significantly reduced, so that such
occurrences will be relatively rare.

In addition, and as described above, the method of the present invention compares
the actual data of the new data unit with the actual data of the previously stored
data unit when a data unit having the same hash value as a previously stored data
unit is detected. As such, the possibility of an erroneous identification of a data
unit as being identical with a data unit actually containing different data, even by
one bit, is eliminated.

Thus, Shnelvar clearly teaches away from this important feature of the claimed invention.

Second, Shnelvar fails to teach the feature of relying on the pseudorandomness of the
digital fingerprint to achieve a pseudorandomness in the physical storage node (e.g., data server
or disk drive) on which a data item is stored. Claim 1 has the following limitation:

wherein the pseudorandom distribution of the digital fingerprints introduces
pseudorandomness into the physical storage nodes at which data items are stored
in the data repository.

Applicant : Norman Margolus et al.
Serial No. : 09/785,535
Filed : February 16, 2001
Page : 28 of 28

Attorney's Docket No.: 11656-002001

The examiner concedes that Shnelvar fails to teach this limitation (Office Action, page 7, lines 17-20). With the claim amended to refer to "physical storage node" rather than merely to "physical location", this difference from Shnelvar is even clearer.

Accordingly, claim 1 is patentable over Shnelvar.

The remaining claims are all properly dependent on claim 1, and thus allowable therewith. Each of the dependent claims adds one or more further limitations that enhance patentability, but those limitations are not presently relied upon. For that reason, and not because applicants agree with the examiner, no rebuttal is offered to the examiner's reasons for rejecting the dependent claims.

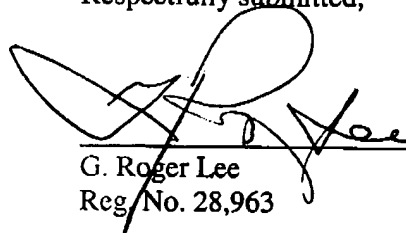
Allowance of the application is requested.

Enclosed is a \$510.00 check for the Petition for Extension of Time fee. Please apply any other charges or credits to deposit account 06-1050.

Respectfully submitted,

Date: _____

3/14/05



G. Roger Lee
Reg. No. 28,963

Fish & Richardson P.C.
225 Franklin Street
Boston, MA 02110-2804
Telephone: (617) 542-5070
Facsimile: (617) 542-8906